


Towards the Realistic, Robust, and Efficient Assessment of Causal Effects with Stochastic Shift Interventions

Considerations for two-phase sampling designs, nonparametric variable importance analysis, and open source software implementations


Nima Hejazi

Friday 14th September, 2018

Group in Biostatistics, and
Center for Computational Biology,
University of California, Berkeley

 nshejazi

 nhejazi

 nimahejazi.org

 bit.ly/2018_biostat_qual



Preview: Summary

- Vaccine efficacy evaluation helps to develop enhanced vaccines better informed by biological properties of the target disease.
- HIV vaccines modulate immune responses as part of the mechanism for lowering HIV risk.
- *Stochastic* interventions provide a flexible framework for considering **realistic** treatment policies.
- Large-scale vaccine trials often use two-phase sampling — need to accommodate such designs.
- We've developed robust, open source statistical software for applying stochastic interventions in observational studies.

This slide deck is for a seminar-length talk (about 50 minutes) on an approach to causal inference and nonparametric variable importance in the context of parameters defined as treatment shifts. Here, we introduce an additive treatment shift parameter, extensions for censored data (including a multiple double robustness property), new opensource statistical software for applying our approach, and applications to a vaccine efficacy trial examining HIV. This talk was most recently given as part of my PhD qualifying examination at the University of California, Berkeley.

Source: <https://github.com/nhejazi/qualexam-phd-biostat>

Slides: http://bit.ly/2018_biostat_qual

With notes: http://bit.ly/2018_biostat_qual_notes

Motivations and Preliminaries

Motivation: The Burden of HIV-1

- HIV-1 epidemic is in its fourth decade, with 2.5 million new infections occurring annually worldwide.
- Though diminishing, number of newly infected persons outpaces number of patients starting antiretroviral therapy.
- As of 2013, progress is limited: the HIV vaccine efficacy trial with highest impact reported a 31% reduction in infections.
- Careful study led to development of a targeted CD4+ and CD8+ T-cell and antibody boost vaccine, which underwent extensive pre-clinical and early-phase clinical testing.

Motivation: The HVTN 505 Trial

- The HIV Vaccine Trials Network (HVTN) 505 preventive vaccine efficacy trial, detailed in Hammer et al. (2013).
- $n = 2504$ participants — all observed cases matched to controls after collection of endpoints of interest.
- Baseline variables (W): sex, age, BMI, SES, etc.
- Variables of interest (A): immune response markers (post-vaccination T-cell activity).
- Outcome of interest (Y): HIV-1 infection status.
- The 505 trial includes both vaccine and placebo arms — we focus on only the vaccine arm (for now).

2

- A vaccine effective at preventing HIV-1 acquisition would be a cost-effective and durable approach to halting the worldwide epidemic.
- Identifying vaccine-induced immune-response biomarkers that predict a vaccine's ability to protect individuals from HIV-1 infection is a high priority.
- The study was halted on 22 April 2013 due to absence of vaccine efficacy. There was no significant effect of the vaccine on the primary infection endpoint of HIV-1 infection between week 28 and month 24.

Interlude: Immune Responses in HVTN 505

- 12-color intracellular cytokine staining (ICS) assay.
- Cryopreserved peripheral blood mononuclear cells were stimulated with synthetic HIV-1 peptide pools.
- Immune responses of interest were
 1. Total magnitude of the CD4⁺ T-cell response.
 2. COMPASS Env-specific CD4⁺ T-cell polyfunctionality.
 3. Total magnitude of the CD8⁺ T-cell response.
 4. COMPASS Env-specific CD8⁺ T-cell polyfunctionality.
 5. CD4⁺ and CD8⁺ T-cell log₁₀-transformed total magnitude.
- All immune responses are assayed *after* the endpoints of interest (HIV-1 infection status) are collected.

3

- For a complete description of the immune responses of interests and how these were collected, consult the supplemental materials of HE Janes (2017).
- This class of data is difficult and expensive to collect, which begins to provide motivation for why it might be undesirable to restrict the types of analyses performed to classical semiparametrics.
- Such classical analyses severely restrict the scope of the scientific questions we're able to ask.

Motivation: Enhanced Vaccines for HIV-1

- **Question:** How would changes in immune response profile (for a given marker) impact HIV-1 infection?
- Using observational data, thought experiments will allow us to examine scenarios where immune responses were tweaked differently by a given vaccine.
- By isolating potential immune response targets, improved vaccines can be constructed with such targets in mind.
- **Conclusion:** Understanding which immune responses impact vaccine efficacy helps develop more efficacious vaccines.

4

- HIV is a high-impact public health issue but numerous attempts to develop vaccines have met with only mild success.
- The complexity of the disease mechanism makes it quite challenging to study the numerous factors that contribute to a possible mitigation of infection risk.

Structure of the Full Data

- Consider $X = (W, A, Y) \sim P_0^X \in \mathcal{M}_{NP}^X$
 - W — baseline covariates (e.g., sex, age, etc.),
 - A — intervention (e.g., immune response profile post-vaccination),
 - Y — outcome of interest (e.g., HIV-1 infection status),
 - P_0^X — true (unknown) distribution of the full data X ,
 - \mathcal{M}_{NP}^X — nonparametric statistical model.
- Consider observing X_1, \dots, X_n , i.e., n iid copies of X .
- For now, ignore two-phase sampling design of HVTN 505.

Likelihood Factorization for the Full Data

- Let $q_{0,Y}$ be the conditional density of Y given (A, W) wrt dominating measure ξ .
- Let $q_{0,A}$ be the conditional density of A given W wrt dominating measure μ .
- Let $q_{0,W}$ be the density of W wrt dominating measure ν .
- Then, for p_0^X , density of X wrt the product measure, density evaluated on a particular observation x :

$$p_0^X(x) = q_{0,Y}^X(y | A = a, W = w)q_{0,A}^X(a | W = w)q_{0,W}^X(w).$$

NPSEM for the Full Data

- Use a nonparametric structural equation model (NPSEM) to describe generation of X (Pearl 2009), specifically

$$W = f_W(U_W)$$

$$A = f_A(W, U_A)$$

$$Y = f_Y(A, W, U_Y)$$

- NPSEM parameterizes p_0^X in terms of the distribution of RVs (X, U) modeled by this system of equations.
- Implies a model for the distribution of counterfactual RVs generated by interventions on the data-generating process.

7

- Notation: let f_W, f_A, f_Y be deterministic functions, and U_W, U_A, U_Y exogenous RVs.

Interlude: On Realism and Causal Inference

- Our motivations, again: How do changes in a given immune response affect post-vaccination risk of HIV-1 infection?
- How does *intervening* on a given immune response *cause* changes in HIV-1 risk? How can future vaccines be informed by the lessons we can learn from such thought experiments?
- What does it mean to intervene on a continuous-valued quantity? Should one consider setting a specific value deterministically? Perhaps with a dynamic intervention?
- Immune responses are surely functions of W , and, perhaps, a posited change should depend on A as well.

8

Classical parameters from causal inference (e.g., ATE) are not well-suited for such questions.

Science × Software × Statistics

Statistical Software Promotes Scientific Reproducibility

“The origins of the scientific method. . . amount to insistence on direct evidence. This is reflected in the motto of The Royal Society, founded in 1660: *Nullius in verba*, which roughly means ‘take nobody’s word for it.’” —Philip B. Stark, in *The Practice of Reproducible Research*

- *Show me, not trust me* — robust, well documented, rigorously tested software allows others to directly assess our claims.
- The software we distribute allows others to engage with our analysis of the HVTN 505 trial and to apply our method to similar vaccine efficacy trials.
- We will be making *causal* claims about vaccine efficacy. Our results could be medically impactful, so reproducibility is vital.

What's wrong with that code I wrote that one time?

“An article. . .in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.” —David Donoho, on *WaveLab*

- Reproducible applied statistics *requires robust software*:
 - Clear, easily accessible documentation, including examples.
 - Unit testing to rigorously assess functions, classes, etc.
 - Continuous integration, ensuring accessibility across systems and constant monitoring of software quality.
 - Open source development, embodying an ongoing, continuous, public peer review of the software product.

- Statistical publications feature simulation and data analytic results, largely insufficient for rigorously assessing claims.
- “*A Year is a Long Time in this Business*. Once, about a year after one of us had done some work and written an article (and basically forgot the details of the work he had done), he had the occasion to apply the methods of the article on a newly-arrived dataset. When he went back to the old software library to try and do it, he couldn’t remember how the software worked — invocation sequences, data structures, etc. In the end, he abandoned the project, saying he just didn’t have time to get in to it anymore.” —from *WaveLab*
- Another example (also from *WaveLab*): this one about how the figures in an interesting paper were produced by software built by the authors, but the authors themselves did not remember the parameter settings that produced the nice figure, so they just hoped that the configuration would come back to them some time later.

Software Ecosystem — The t1verse!

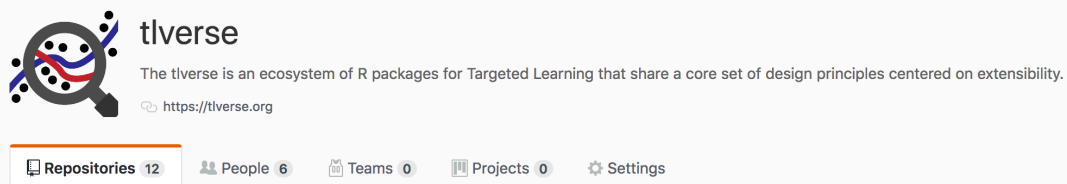


Figure 1: <https://github.com/t1verse>

- A new software environment and framework for Targeted Learning, with a focus on extensibility.
- *Goal:* expose a set of design principles off of which all future Targeted Learning software will be built.
- *How?* A core framework, plus minimal extensions (“connector packages”), each implementing a few estimators.

- Contribute on GitHub: <https://github.com/tlverse>.
- Reach out to us with questions and any feature requests.

Towards Stochastic Interventions

Static or Dynamic Interventions in our NPSEM

- A *static intervention* replaces f_A with a specific value a in the support of $q_{0,A}(\cdot | W)$ a.e. $q_{0,W}$.
- This requires specifying a particular value of the intervention (i.e., $A = a$) under which to evaluate the outcome.
- Is it really sensible to set the value of an immune response to a particular value? (Even as a function of W .)
- This is statistically accessible but scientifically uninteresting (perhaps dishonest, even) — one would have to test many different values a , with only intuition as a guide.

12

- We'd like to posit changes to immune responses that we can actually induce.
- We can mathematically consider the counterfactual of setting an immune response to a specific value, but we cannot do this in practice — thus, it is rather uninteresting.
- Instead, we consider counterfactuals that might come to pass. We are able to create vaccines that alter the level of an immune response by a small amount.

Interlude: A Linear Modeling Perspective

- Briefly consider a simple data structure: $X = (Y, A)$; we seek to model the outcome Y as a function of A .
- To posit a linear model, consider $Y_i = \beta_0 + \beta_1 A_i + \epsilon_i$, with error $\epsilon_i \sim N(0, 1)$.
- Letting δ be a change in A , $Y_{A+\delta} - Y_A$ may be expressed

$$\begin{aligned}\mathbb{E}Y_{A+\delta} - \mathbb{E}Y_A &= [\beta_0 + \beta_1(\mathbb{E}A + \delta)] - [\beta_0 + \beta_1(\mathbb{E}A)] \\ &= \beta_0 - \beta_0 + \beta_1\mathbb{E}A - \beta_1\mathbb{E}A + \beta_1\delta \\ &= \beta_1\delta\end{aligned}$$

- Thus, a *unit shift* in A (i.e., $\delta = 1$) may be seen as inducing a change in the difference in outcomes of magnitude β_1 .

13

- We extend this result to the mean counterfactual outcomes under the nonparametric model \mathcal{M} .
- Linear modeling analogy re: conversation with Alan on 22 August.

Interlude: A Causal Inference Perspective

- Consider a data structure: $(Y_a, a \in \mathcal{A})$.
- To posit a linear model, let $Y_a = \beta_0 + \beta_1 a + \epsilon_a$ for $a \in \mathcal{A}$, with error $\epsilon_a \sim N(0, \sigma_a^2) \forall a \in \mathcal{A}$.
- For the counterfactual outcomes $(Y_{a'+\delta}, Y_{a'})$, their difference, $Y_{a'+\delta} - Y_{a'}$, for some $a' \in \mathcal{A}$, may be expressed

$$\begin{aligned}\mathbb{E}Y_{a'+\delta} - \mathbb{E}Y_{a'} &= [\beta_0 + \beta_1(a' + \delta) + \mathbb{E}\epsilon_{a'+\delta}] - [\beta_0 + \beta_1 a' + \mathbb{E}\epsilon_{a'}] \\ &= \beta_1 \delta\end{aligned}$$

- Thus, a *unit shift* for $a' \in \mathcal{A}$ (i.e., $\delta = 1$) may be seen as inducing a change in the difference in the counterfactual outcomes of magnitude β_1 .

14

- Note that this analysis is exactly what we're told we **cannot** do in linear models 101 — that is, the slope of a regression line cannot be interpreted as *causing* a change in the outcome.
- We extend this result to the mean counterfactual outcomes under the nonparametric model \mathcal{M} .
- Linear modeling analogy re: conversation with Alan on 22 August.
- Example updated to incorporate counterfactuals re: conversation with David on 30 August

Stochastic Interventions in our NPSEM

- *Stochastic interventions* modify the value A would naturally assume, $f_A(W, U_A)$, by drawing from a modified intervention distribution $G^*(\cdot | W)$ so that the new value $A^* \sim G^*(\cdot | W)$.
- This generates a counterfactual RV, with distribution P_0^d , $Y_{d(A,W)} := f_Y(d(A, W), W, U_Y) \equiv Y_{G^*} := f_Y(A^*, W, U_Y)$.
- We estimate $\psi_{0,d} := \mathbb{E}_{P_0^d}\{Y_{d(A,W)}\}$, mean of $Y_{d(A,W)}$.
- For HVTN 505, $\psi_{0,d}$ is the counterfactual risk of HIV-1 infection, had the observed value of the immune response been altered under the rule $d(A, W)$ defining $G^*(\cdot | W)$.
- Helps isolate immune responses that causally inhibit HIV-1; future vaccines may be designed to target these markers.

15

- Several different ways to consider stochastic interventions.
- Starts with Mark and Ivan's simple stochastic shift.
- Extensions to modified treatment policies.
- The new value of A may be denoted $A^* \sim G^*(\cdot | W)$, where $A^* = d(W, U^*)$ for a rule d and random error U^* .

Literature: Díaz and van der Laan (2012)

- *Proposal*: Evaluate outcome under an altered *intervention distribution* — e.g., $P_\delta(g_0)(A = a | W) = g_0(a - \delta(W) | W)$.
- Identification conditions for a statistical parameter of the counterfactual outcome $\psi_{0,d}$ under such an intervention.
- Show that the causal quantity of interest $\mathbb{E}_0\{Y_{d(A,W)}\}$ is identified by a functional of the distribution of X :

$$\psi_{0,d} = \int_{\mathcal{W}} \int_{\mathcal{A}} \mathbb{E}_{P_0^X}\{Y | A = d(a, w), W = w\} \cdot q_{0,A}^X(a | W = w) \cdot q_{0,W}^X(w) d\mu(a) d\nu(w)$$

- Provides a derivation based on the efficient influence function (EIF) with respect to the nonparametric model \mathcal{M} .

16

- The identification result allows us to write down the causal quantity of interest in terms of a functional of the observed data.
- Key innovation: loosening standard assumptions through a change in the observed intervention mechanism.
- Problem: globally altering an intervention mechanism does not necessarily respect individual characteristics.
- The authors build IPW, A-IPW, and TML estimators, comparing the three different approaches.
- **IMPORTANT**: gives the g-computation formula for identification of this estimator from the observed data structure.

Literature: Haneuse and Rotnitzky (2013)

- *Proposal*: Characterization of stochastic interventions as *modified treatment policies* (MTPs).
- Assumption of *piecewise smooth invertibility* allows for the intervention distribution of any MTP to be recovered:

$$g_{0,\delta}(a | w) = \sum_{j=1}^{J(w)} I_{\delta,j}\{h_j(a, w), w\} g_0\{h_j(a, w) | w\} h'_j(a, w)$$

- Such intervention policies account for the natural value of the intervention A directly yet are interpretable as the imposition of an altered intervention mechanism.
- Identification conditions for assessing the parameter of interest under such interventions appear technically complex (at first).

17

- Shifts of the form $d(A, W)$ are considerably more interesting since these are realistic intervention policies.
- Example: consider an individual with an extremely high immune response but whose baseline covariates W suggest we shift the response still higher. Such a shift may not be biologically plausible (impossible, even) but we cannot account for this if the shift is only a function of W .
- The authors build IPW, outcome regression, and non-iterative doubly robust estimators, as well as an approach based on MSMs.
- Piecewise smooth invertibility: This assumption ensures that we can use the change of variable formula when computing integrals over A and it is useful to study the estimators that we propose in this paper.

Literature: Díaz and van der Laan (2018)

- Builds on the original proposal, accommodating MTP-type shifts $d(A, W)$ proposed after their earlier work.
- To protect against positivity violations, considers a specific shifting mechanism:

$$d(a, w) = \begin{cases} a + \delta, & a + \delta < u(w) \\ a, & \text{otherwise} \end{cases}$$

- Proposes an improved “1-TMLE” algorithm, with a single auxiliary covariate for constructing the TML estimator.
- Our (first) contribution: implementation of this algorithm.

From the Causal to the Statistical Target Parameter

Assumption 1: *Consistency*

$Y_i^{d(a_i, w_i)} = Y_i$ in the event $A_i = d(a_i, w_i)$, for $i = 1, \dots, n$

Assumption 2: *SUTVA*

$Y_i^{d(a_i, w_i)}$ does not depend on $d(a_j, w_j)$ for $i = 1, \dots, n$ and $j \neq i$, or lack of interference (Rubin 1978; 1980)

Assumption 3: *Strong ignorability*

$A_i \perp\!\!\!\perp Y_i^{d(a_i, w_i)} \mid W_i$, for $i = 1, \dots, n$

Assumption 4: Positivity (or overlap)

$a_i \in \mathcal{A} \implies d(a_i, w_i) \in \mathcal{A}$ for all $w \in \mathcal{W}$, where \mathcal{A} denotes the support of A conditional on $W = w_i$ for all $i = 1, \dots, n$

- This positivity assumption is not quite the same as that required for categorical interventions.
- In particular, we do not require that the intervention density place mass across all strata defined by W .
- Rather, we merely require the post-intervention quantity be seen in the observed data for given $a_i \in \mathcal{A}$ and $w_i \in \mathcal{W}$.

Statistical Target Parameter for Shift Interventions

- Now we have the statistical functional (target parameter):

$$\Psi(P_0^X) = \mathbb{E}_{P_0^X} \bar{Q}(d(A, W), W),$$

allowing estimation of causal parameter $\psi_{0,d} = \mathbb{E} Y_{d(A,W)}$.

- We now seek to efficiently estimate this target parameter in the nonparametric model \mathcal{M} .
- We implement a targeted minimum loss-based estimator of this statistical target parameter.

21

- Note that $\bar{Q}(\cdot | W)$ is properly referred to as a *functional of the likelihood*.

Interlude: Slope in a Semiparametric Model

- Consider the stochastic intervention $g^*(\cdot | W)$:

$$\begin{aligned}\mathbb{E}Y_{g^*} &= \int_W \int_a \mathbb{E}(Y | A = a, W) g(a - \delta | W) \cdot da \cdot dP_0(W) \\ &= \int_W \int_z \mathbb{E}(Y | A = z + \delta, W) g(z | W) \cdot dz \cdot dP_0(W),\end{aligned}$$

defining the change of variable $z = a - \delta$.

- For a semiparametric model, $\mathbb{E}(Y | A = z, W) = \beta z + \theta(W)$:

$$\begin{aligned}\mathbb{E}Y_{g^*} - \mathbb{E}Y &= \int_W \int_z [\mathbb{E}(Y | A = z + \delta, W) - \mathbb{E}(Y | A = z, W)] \\ &\quad g(z | W) \cdot dz \cdot dP_0(W) \\ &= [\beta(z + \delta) + \theta(W)] - [\beta z + \theta(W)] \\ &= \beta \delta\end{aligned}$$

Targeted Minimum Loss-Based Estimation

- A TMLE is an algorithm for updating initial estimators so as to satisfy an arbitrary set of estimating equations.
- Semiparametric-efficient estimation thru solving efficient influence function estimating equation wrt the model \mathcal{M} .
- Statistical target parameter: $\Psi(P_0^X) = \mathbb{E}_{P_0^X} \bar{Q}(d(A, W), W)$
- For which the efficient influence function (EIF) is

$$D(P_0^X)(x) = H(a, w)(y - \bar{Q}(a, w)) + \bar{Q}(d(a, w), w) - \Psi(P_0^X)$$

- The auxiliary covariate $H(a, w)$ may be expressed

$$H(a, w) = \mathbb{I}(a + \delta < u(w)) \frac{g_0(a - \delta | w)}{g_0(a | w)} + \mathbb{I}(a + \delta \geq u(w))$$

23

- The auxiliary covariate simplifies when the treatment is in the limits (conditional on W) — i.e., for $A_i \in (u(w) - \delta, u(w))$, then we have $H(a, w) = \frac{g_0(a - \delta | w)}{g_0(a | w)} + 1$.
- Need to explicitly remind the audience what $u(w)$ is again. It's only appeared once at this point, and only been mentioned in passing.

Key Properties of TML Estimators

- **Asymptotic linearity:**

$$\Psi(P_n^*) - \Psi(P_0^X) = \frac{1}{n} \sum_{i=1}^n D(P_0^X)(X_i) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

- **Gaussian limiting distribution:**

$$\sqrt{n}(\Psi(P_n^*) - \Psi(P_0^X)) \rightarrow N(0, \text{Var}(D(P_0^X)(X)))$$

- **Statistical inference:**

$$\text{Wald-type CI : } \Psi(P_n^*) \pm z_\alpha \cdot \frac{\sigma_n}{\sqrt{n}},$$

where σ_n^2 is computed directly via $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n D^2(\cdot)(X_i)$.

24

Under the additional condition that the remainder term $R(\hat{P}^*, P_0)$ decays as $o_P\left(\frac{1}{\sqrt{n}}\right)$, we have that $\Psi_n - \Psi_0 = (P_n - P_0) \cdot D(P_0) + o_P\left(\frac{1}{\sqrt{n}}\right)$, which, by a central limit theorem, establishes a Gaussian limiting distribution for the estimator, with variance $V(D(P_0))$, the variance of the efficient influence function when Ψ admits an asymptotically linear representation.

The above implies that Ψ_n is a \sqrt{n} -consistent estimator of Ψ , that it is asymptotically normal (as given above), and that it is locally efficient. This allows us to build Wald-type confidence intervals, where σ_n^2 is an estimator of $V(D(P_0))$. The estimator σ_n^2 may be obtained using the bootstrap or computed directly via $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n D^2(\bar{Q}_n^*, g_n)(O_i)$

We obtain semiparametric-efficient estimation and robust inference in the nonparametric model \mathcal{M} by solving the efficient influence function.

1. If $D(\bar{Q}_n^*, g_n)$ converges to $D(P_0)$ in $L_2(P_0)$ norm.
2. The size of the class of functions \bar{Q}_n^* and g_n is bounded (technically, $\exists \mathcal{F}$ st $D(\bar{Q}_n^*, g_n) \in \mathcal{F}$ whp, where \mathcal{F} is a Donsker class)

Algorithm for TML Estimation

1. Construct initial estimators g_n of $g_0(A, W)$ and Q_n of $\bar{Q}_0(A, W)$, perhaps using data-adaptive regression techniques.
2. For each observation i , compute an estimate $H_n(a_i, w_i)$ of the auxiliary covariate $H(a_i, w_i)$.

3. Estimate the parameter ϵ in the logistic regression model

$$\text{logit} \bar{Q}_{\epsilon, n}(a, w) = \text{logit} \bar{Q}_n(a, w) + \epsilon H_n(a, w),$$

or an alternative regression model incorporating weights.

4. Compute TML estimator Ψ_n of the target parameter, defining update \bar{Q}_n^* of the initial estimate \bar{Q}_{n, ϵ_n} :

$$\Psi_n = \Psi(P_n^*) = \frac{1}{n} \sum_{i=1}^n \bar{Q}_n^*(d(A_i, W_i), W_i).$$

25

- We recommend using nonparametric methods for the initial estimators, as consistent estimation is necessary for efficiency of the estimator Ψ_n .
- Intuition for the submodel fluctuation?

R/ s13 : modern Super Learning with pipelines

build passing build passing coverage 82% repo status Active License GPL v3 DOI 10.5281/zenodo.1342294 chat on gitter

A modern implementation of the Super Learner algorithm for ensemble learning and model stacking

Authors: [Jeremy Coyle](#), [Nima Hejazi](#), [Ivana Malenica](#), [Oleg Sofrygin](#)

What's s13 ?

s13 is a modern implementation of the Super Learner algorithm of van der Laan, Polley, and Hubbard (2007). The Super

Figure 2: <https://github.com/tlverse/s13>

- A robust and efficient implementation of the Super Learner algorithm, drawing on the concept of pipelines popularized by `scikit-learn` (<http://scikit-learn.org/stable/>).
- The first software package and one of the core engines of the `tlverse` ecosystem.

26

- Contribute on GitHub: <https://github.com/tlverse/s13>.
- Reach out to us with questions and any feature requests.

Simulation Study: TML Estimation

- For a single observational unit $X = (W, A, Y)$, data are simulated using the following set of structural equations:

$$W \sim \text{Bern}(p = 0.5)$$

$$A \sim N(\mu = \gamma \cdot W, \sigma^2 = 1)$$

$$Y = A + W + \epsilon,$$

- Let $\gamma = 2$ be a multiplier of the effect of the baseline covariate W on the natural value of the treatment A , and white noise $\epsilon \sim N(0, 1)$.
- We consider the case of observing a data structure composed of n replicates of X , i.e., X_1, \dots, X_n .
- Letting $\delta = 0.5$, we construct a TML estimate of the counterfactual mean outcome $\psi_{0,d}$.

27

- Sample sizes $n_{\text{samp}} \in \{50, 250, 1000, 2000\}$, each with $n_{\text{sim}} = 1000$ simulations were performed at each sample size.

Simulation Study: TML Estimation

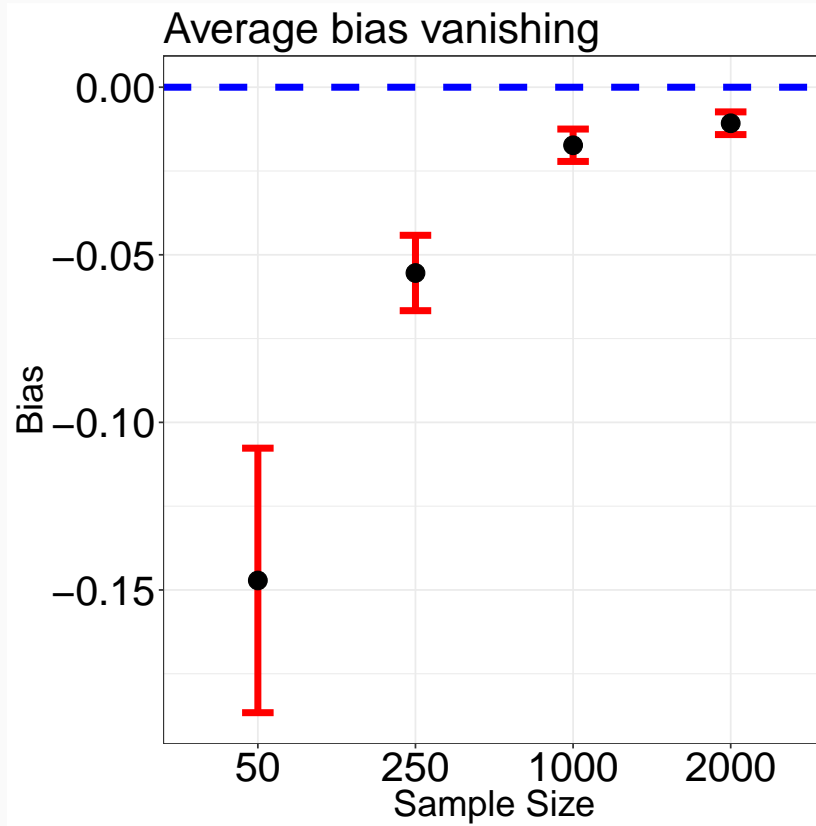
n_{samp}	n_{sim}	Bias	MC Var	SE	MSE	Coverage
50	753	-0.1471	0.3052	0.3104	0.3269	0.6999
250	845	-0.0554	0.0278	0.1361	0.0309	0.8615
1000	922	-0.0173	0.0056	0.0672	0.0059	0.9284
2000	946	-0.0107	0.0028	0.0475	0.0029	0.9070

Table 1: Average estimates of properties of the TML estimator.

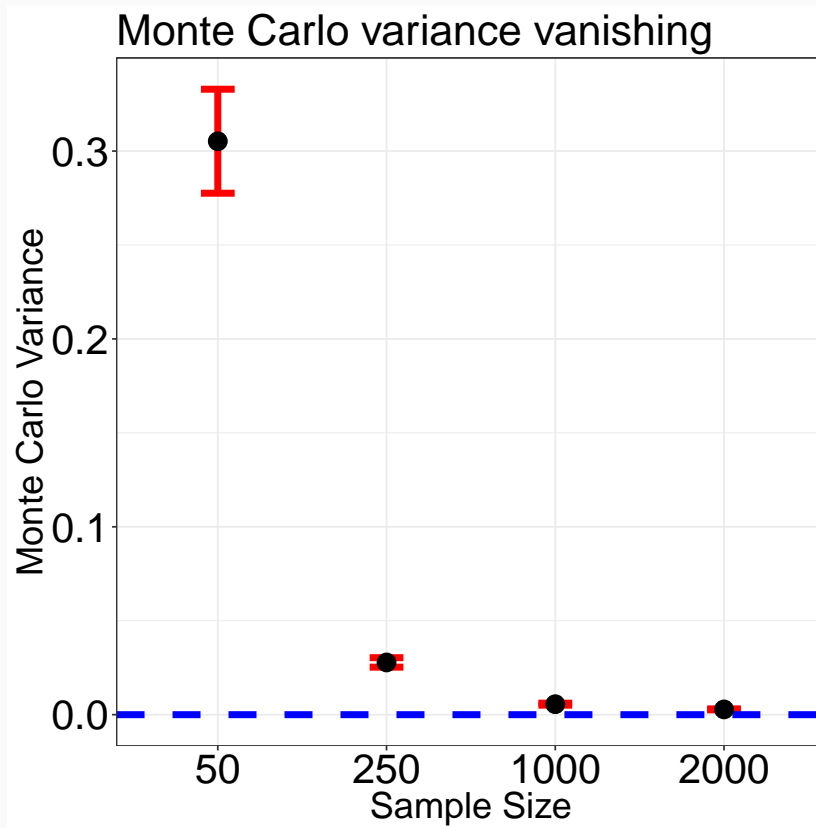
n_{samp}	n_{sim}	Bias	MC Var	MSE
50	753	0.0395	0.0277	0.0310
250	845	0.0112	0.0025	0.0029
1000	922	0.0048	0.0005	0.0006
2000	946	0.0034	0.0003	0.0003

Table 2: Error in estimates of properties of the TML estimator.

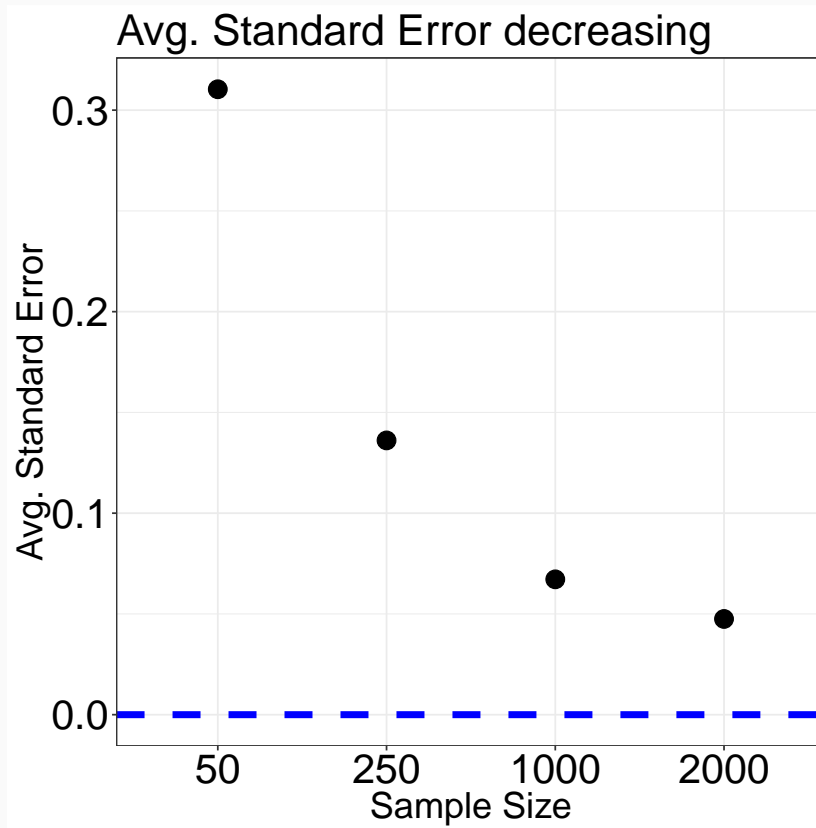
Simulation Study: TML Estimation



Simulation Study: TML Estimation



Simulation Study: TML Estimation



R/ tmle3shift

build passing build passing coverage 93% repo status Active License GPL v3

Targeted Learning and Variable Importance with Stochastic Interventions

Authors: [Nima Hejazi](#), [Jeremy Coyle](#), and [Mark van der Laan](#)

What's tmle3shift ?

tmle3shift is an adapter/extension R package in the `tlverse` ecosystem that exposes support for the estimation of a target

Figure 3: <https://github.com/tlverse/tmle3shift>

- Tools for assessing the effects of stochastic interventions.
- Supports interventions that enforce positivity constraints.
- First of many “connector” R packages that extend the `tlverse` ecosystem.

32

- Contribute on GitHub:
<https://github.com/tlverse/tmle3shift>.
- Reach out to us with questions and any feature requests.

Two-Phase Sampling Designs

Data Structure for Two-Phase Designs

- In the 505 HIV-1 trial, all infected individuals are matched to controls after endpoints are collected.
- We need to extend our full data structure $X = (W, A, Y)$ to accommodate such a sampling procedure.
- Consider the observed data structure $O = (W, \Delta, \Delta A, Y)$, a masked version of the full data structure.
- Let $\Delta = f(Y, W)$ be binary s.t. $\Delta \in \{0, 1\}$, where $\Delta = 1$ corresponds to being selected into the second-stage sample.
- Let $\pi_0(Y, W) = \mathbb{P}(\Delta = 1 \mid Y, W)$, and let $\pi_n(Y, W)$ be an estimator of $\pi_0(Y, W)$.

- Note that matching is done based on $\{W, Y\}$.

Augmented Estimators for Two-Phase Designs

- Rose and van der Laan (2011) introduce the IPCW-TMLE, to be used when observed data is subject to two-phase sampling.
- Their proposal constructs estimators for an observed data structure of the form $O = (V, \Delta, \Delta X)$.
- In our use-case, the sampling node $V = \{Y, W\}$, and thus we have our proposed data structure $O = (W, \Delta, \Delta A, Y)$.
- *Initial proposal*: correct for two-phase sampling by using an IPC-weighted loss function:

$$\mathcal{L}(P_0^X)(O) = \frac{\Delta}{\pi_n(Y, W)} \mathcal{L}^F(P_0^X)(X)$$

Efficiency Under Two-Phase Sampling

- When the sampling mechanism is not known by design, it is best to employ a nonparametric estimator of $\pi_0(Y, W)$.
- When $\pi_0(Y, W)$ is estimated nonparametrically, the IPCW augmentation must be applied to the EIF:

$$D(P_0^X)(o) = \frac{\Delta}{\pi_0(y, w)} D^F(P_0^X)(x) - \left(1 - \frac{\Delta}{\pi_0(y, w)}\right) \mathbb{E}(D^F(P_0^X)(x) \mid \Delta = 1, Y = y, W = w),$$

expressed in terms of the full data EIF $D^F(P_0^X)(x)$.

Efficiency Under Two-Phase Sampling

The IPC-augmented EIF points out two distinct terms:

$$\frac{\Delta}{\pi_0(y,w)} D^F(P_0^X)(x)$$

The IPC-weighted EIF of the full data structure X , relative to the nonparametric model \mathcal{M} ; and,

$$\left(1 - \frac{\Delta}{\pi_0(y,w)}\right) \mathbb{E}(D^F(P_0^X)(x) \mid \Delta = 1, Y = y, W = w)$$

The expectation of the full data EIF $D^F(P_0^X)(x)$, taken only over units selected by the sampling mechanism (i.e., $\Delta = 1$).

Emergent Property: Multiple Robustness

- We now have a semiparametric-efficient and robust procedure for assessing the effect of the intervention $d(a, w) = a + \delta$.
- Due to the construction of the IPCW-TMLE, the resultant estimator is robust and efficient under two-phase sampling.
- Uniquely, a multiple robustness property emerges — through combinations of (g, Q) and $(\pi(Y, W), \mathbb{E}(D^F(P_0^X)(x) | Y, W))$.
- This allows us to assess how posited shifts in the assayed immune responses would have affected HIV-1 infection risk.

Simulation Study: IPC-Weighted TML Estimation

- For a single observational unit $O = (W, \Delta, \Delta A, Y)$, data are simulated using the following set of structural equations:

$$W_1 \sim N(\mu = 3, \sigma^2 = 1)$$

$$W_2 \sim \text{Bern}(p = 0.6)$$

$$W_3 \sim \text{Bern}(p = 0.3)$$

$$A \sim N(\mu = 2 \cdot (W_2 + W_3), \sigma^2 = 1)$$

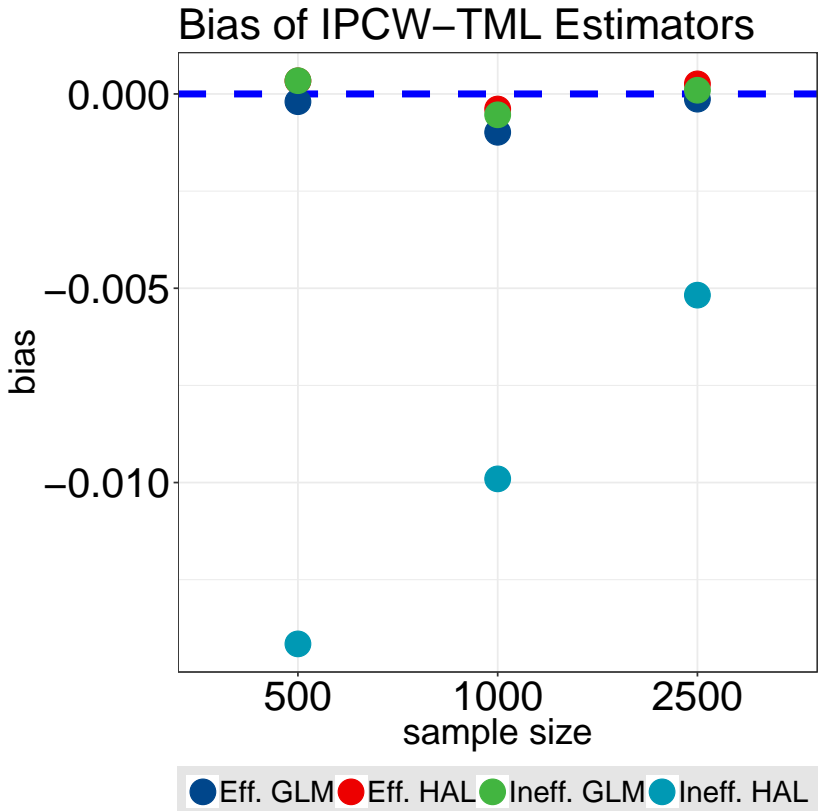
$$Y = \text{Bern} \left(p = \frac{\left(1 + \tanh \left(\frac{W_1 + W_2 + W_3 - A}{3} \right) \right)}{2} \right)$$

$$\Delta = \text{Bern} \left(p = \frac{\left(1 + \tanh \left(\frac{W_1 + W_2 + W_3 - Y}{3} \right) \right)}{2} \right)$$

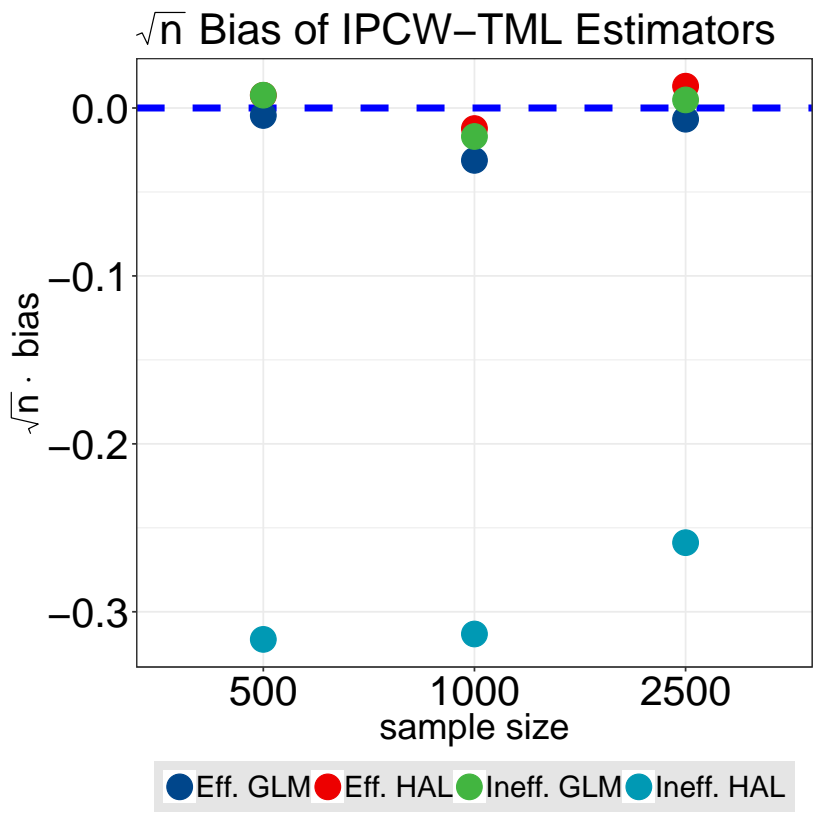
Simulation Study: IPC-Weighted TML Estimation

- We consider the case of observing a data structure composed of n replicates of O , i.e., O_1, \dots, O_n .
- Letting $\delta = 0.5$, we construct an IPCW-TML estimate of the counterfactual mean outcome $\psi_{0,d}$ for P_0^X , the data generating distribution of the full data X from which O is derived.
- **Goal:** Assess extent to which fitting sampling mechanism with a nonparametric regression affects the resultant estimator.
 1. Fit $\pi_0(Y, W)$ with a GLM or the Highly Adaptive Lasso (HAL), building loss-augmented and EIF-augmented TMLEs.
 2. Compare bias, variance, and relative efficiency of the resultant TML estimators.

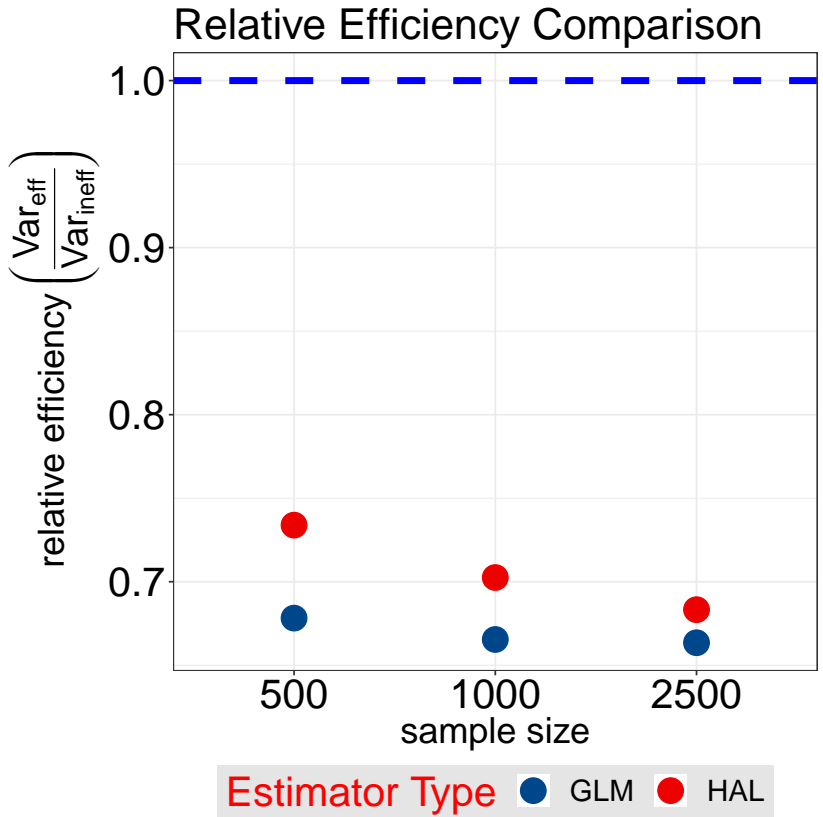
Simulation Study: IPC-Weighted TML Estimation



Simulation Study: IPC-Weighted TML Estimation



Simulation Study: IPC-Weighted TML Estimation



Software package: R/txshift

R/ txshift

build passing build passing coverage 47% repo status Active license MIT

Targeted Learning of the Causal Effects of Stochastic Interventions

Authors: [Nima Hejazi](#) and [David Benkeser](#)

What's txshift ?

The txshift R package is designed to provide facilities to compute targeted maximum likelihood estimates (TMLE) of the

Figure 4: <https://github.com/nhejazi/txshift>

- Supports estimation of the effects of simple (additive) stochastic interventions.
- Implements both types of IPCW-TML estimator, allowing for two-phase sampling to be appropriately handled when $\pi_0(V)$ is known by design or unknown.

- Contribute on GitHub: <https://github.com/nhejazi/txshift>.
- Reach out to us with questions and any feature requests.

Extensions and Future Directions

Ongoing Efforts

- Extensions of stochastic interventions to causal mediation analysis — new theory provides estimators of the *natural direct effect* and the *natural indirect effect*.
 - Collaboration with Iván Díaz (Cornell) in progress.
- Further refinement of the t1verse software ecosystem, including new “connector” R packages.
- Data analysis of the HVTN 505 HIV-1 vaccine trial, and discussion of the scientific findings with scientist collaborators.

Future Directions

- Exploration of different forms of stochastic interventions — Kennedy (2018) proposes a shift in propensity scores for binary (or categorical) interventions.
 - Implementation in the `tlverse` ecosystem.
- Refinements of statistical theory so as to better work with quantities common in survival analysis: hazards? survival?
- Assessment of newly concluded and ongoing efficacy trials through work with ongoing collaborators at Fred Hutch.

Software and Statistics Revisited

software	data	CI	testing	docs
tmle3shift	(W, A, Y)	✓	✓	✓
txshift	(W, A, Δ, Y)	✓	✓	✓
medshift	(W, A, Z, Y)	IP	IP	IP
tmle3	(W, A, Y)	✓	✓	✓
sl3	(X, Y)	✓	✓	✓

- Software is **not** an ancillary activity: How can a theorem or result impact science if no one can apply it?
- Writing software is important for learning statistics:
 - How do people plan to use the software?
 - What is the problem that the software solves?
 - What's the "best" way to estimate the quantity of interest?
- Writing software *impacts* statistics — minor tweaks to implemented estimators help us discover new ideas.

46

- How I spend my time: 55% software, 25% writing, 10% reading, 10% simulations, 5% data analysis.
- Other R packages: biotmle, methyvim, survtmle, survsl (IP), origami, hal9001

Review: Summary

- Vaccine efficacy evaluation helps to develop enhanced vaccines better informed by biological properties of the target disease.
- HIV vaccines modulate immune responses as part of the mechanism for lowering HIV risk.
- *Stochastic* interventions provide a flexible framework for considering **realistic** treatment policies.
- Large-scale vaccine trials often use two-phase sampling — need to accommodate such designs.
- We've developed robust, open source statistical software for applying stochastic interventions in observational studies.

47

It's always good to include a summary.

References

- Breiman, L. (1996). Stacked regressions. *Machine Learning*, 24(1):49–64.
- Díaz, I. and van der Laan, M. J. (2011). Super learner based conditional density estimation with application to marginal structural models. *The international journal of biostatistics*, 7(1):1–20.
- Díaz, I. and van der Laan, M. J. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics*, 68(2):541–549.
- Díaz, I. and van der Laan, M. J. (2018). Stochastic treatment regimes. In *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*, pages 167–180. Springer Science & Business Media.

48

- Dudoit, S. and van der Laan, M. J. (2005). Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology*, 2(2):131–154.
- Hammer, S. M., Sobieszczyk, M. E., Janes, H., Karuna, S. T., Mulligan, M. J., Grove, D., Koblin, B. A., Buchbinder, S. P., Keefer, M. C., Tomaras, G. D., et al. (2013). Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine. *New England Journal of Medicine*, 369(22):2083–2092.
- Haneuse, S. and Rotnitzky, A. (2013). Estimation of the effect of interventions that modify the received treatment. *Statistics in medicine*, 32(30):5260–5277.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960.

49

- Janes, H. E., Cohen, K. W., Frahm, N., De Rosa, S. C., Sanchez, B., Hural, J., Magaret, C. A., Karuna, S., Bentley, C., Gottardo, R., et al. (2017). Higher t-cell responses induced by dna/rad5 hiv-1 preventive vaccine are associated with lower hiv-1 infection risk in an efficacy trial. *The Journal of infectious diseases*, 215(9):1376–1385.
- Kennedy, E. H. (2018). Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, (just-accepted).
- Kennedy, E. H., Ma, Z., McHugh, M. D., and Small, D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1229–1245.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

- Rose, S. and van der Laan, M. J. (2011). A targeted maximum likelihood estimator for two-stage designs. *The International Journal of Biostatistics*, 7(1):1–21.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, pages 34–58.
- Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331.

- van der Laan, M. J., Dudoit, S., and Keles, S. (2004). Asymptotic optimality of likelihood-based cross-validation. *Statistical Applications in Genetics and Molecular Biology*, 3(1):1–23.
- van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super Learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1).
- van der Laan, M. J. and Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1).
- Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2):241–259.
- Young, J. G., Hernán, M. A., and Robins, J. M. (2014). Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic methods*, 3(1):1–19.

Acknowledgments

Close Collaborators and Advisors:


Mark J. van der Laan	University of California, Berkeley
Alan E. Hubbard	University of California, Berkeley
David C. Benkeser	Emory University
Jeremy R. Coyle	University of California, Berkeley
Peter B. Gilbert	Fred Hutchinson Cancer Research Center
Holly E. Janes	Fred Hutchinson Cancer Research Center

Thank you.

Slides: bit.ly/2018_biostat_qual



Notes: bit.ly/2018_biostat_qual_notes

 <https://nimahejazi.org>

 <https://github.com/nhejazi>

 <https://twitter.com/nshejazi>

54

Appendix

Literature: Young et al. (2014)

- Establishes equivalence between g-formula when proposed intervention depends on natural value and when it does not.
- This equivalence leads to a sufficient positivity condition for estimating the counterfactual mean under MTPs via the same statistical functional studied in Díaz and van der Laan (2012).
- Extends earlier identification results, providing a way to use the same statistical functional to assess $\mathbb{E}Y_{d(A,W)}$ or $\mathbb{E}Y_{d(W)}$.
- The authors also consider limits on implementing shifts $d(A, W)$, and address working in a longitudinal setting.

Nonparametric Conditional Density Estimation

- To compute the auxiliary covariate $H(a, w)$, we need to estimate conditional densities $g(A | W)$ and $g(A - \delta | W)$.
- There is a rich literature on density estimation, we follow the approach proposed in Díaz and van der Laan (2011).
- To build a conditional density estimator, consider

$$g_{n,\alpha}(a | W) = \frac{\mathbb{P}(A \in [\alpha_{t-1}, \alpha_t) | W)}{\alpha_t - \alpha_{t-1}},$$

for $\alpha_{t-1} \leq a < \alpha_t$.

- This is a classification problem, where we estimate the probability that a value of A falls in a bin $[\alpha_{t-1}, \alpha_t)$.
- The choice of the tuning parameter t corresponds roughly to the choice of bandwidth in classical kernel density estimation.

Nonparametric Conditional Density Estimation

- Díaz and van der Laan (2011) propose a re-formulation of this classification approach as a set of hazard regressions.
- To effectively employ this proposed re-formulation, consider

$$\mathbb{P}(A \in [\alpha_{t-1}, \alpha_t) \mid W) = \mathbb{P}(A \in [\alpha_{t-1}, \alpha_t) \mid A \geq \alpha_{t-1}, W) \times \prod_{j=1}^{t-1} \{1 - \mathbb{P}(A \in [\alpha_{j-1}, \alpha_j) \mid A \geq \alpha_{j-1}, W)\}$$

- The likelihood of this model may be expressed to correspond to the likelihood of a binary variable in a data set expressed via a long-form repeated measures structure.
- Specifically, the observation of X_i is repeated as many times as intervals $[\alpha_{t-1}, \alpha_t)$ are before the interval to which A_i belongs, and the binary variables indicating $A_i \in [\alpha_{t-1}, \alpha_t)$ are recorded.

Density Estimation with the Super Learner Algorithm

- To estimate $g(A | W)$ and $g(A - \delta | W)$, use a pooled hazard regression, spanning the support of A .
 - We rely on the Super Learner algorithm of van der Laan et al. (2007) to build an ensemble learner that optimally weights each of the proposed regressions, using cross-validation (CV).
 - The Super Learner algorithm uses V -fold CV to train each proposed regression model, weighting each by the inverse of its average risk across all V holdout sets.
 - By using a library of regression estimators, we invoke the result of van der Laan et al. (2004), who prove this likelihood-based cross-validated estimator to be asymptotically optimal.
-
- The auxiliary covariate simplifies when the treatment is in the limits (conditional on W) — i.e., for $A_i \in (u(w) - \delta, u(w))$, then we have $H(a, w) = \frac{g_0(a-\delta|w)}{g_0(a|w)} + 1$.
 - Asymptotically optimal in the sense that it performs as well as the oracle selector as the sample size increases.

Algorithm for IPCW-TML Estimation

1. Using all observed units (X), estimate sampling mechanism $\pi(Y, W)$, perhaps using data-adaptive regression methods.
2. Using only observed units in the second-stage sample $\Delta = 1$, construct initial estimators $g_n(A, W)$ and $\bar{Q}_n(A, W)$, weighting by the sampling mechanism estimate $\pi_n(Y, W)$.
3. With the approach described for the full data case, compute $H_n(a_i, w_i)$, and fluctuate submodel via logistic regression.
4. Compute IPCW-TML estimator Ψ_n of the target parameter, by solving the IPCW-augmented EIF estimating equation.
5. Iteratively update estimated sampling weights $\pi_n(Y, W)$ and IPCW-augmented EIF, updating TML estimate in each iteration, until $\frac{1}{n} \sum_{i=1}^n \text{EIF}_i < \frac{1}{n}$.

- We recommend using nonparametric methods for the initial estimators, as consistent estimation is necessary for efficiency of the estimator Ψ_n .
- Intuition for the submodel fluctuation?
- This process includes the use of HAL to fit the regression of the EIF contributions on the sampling node $\{Y, W\}$.

A Realistic Shift Intervention

Consider a more sophisticated shift function:

$$\delta(a, w) = \begin{cases} \delta, & \delta_{\min}(a, w) \leq \delta \leq \delta_{\max}(a, w) \\ \delta_{\max}(a, w), & \delta \geq \delta_{\max}(a, w) \\ \delta_{\min}(a, w), & \delta \leq \delta_{\min}(a, w) \end{cases},$$

where we define maximal and minimal possible shifts:

$$\delta_{\max}(a, w) = \operatorname{argmax}_{\left\{ \delta \geq 0, \frac{g(a-\delta|w)}{g(a|w)} \leq M \right\}} \frac{g(a-\delta|w)}{g(a|w)}$$

and

$$\delta_{\min}(a, w) = \operatorname{argmin}_{\left\{ \delta \leq 0, \frac{g(a-\delta|w)}{g(a|w)} \leq M \right\}} \frac{g(a-\delta|w)}{g(a|w)}.$$

Variable Importance Analysis with MSMs

- Consider now a grid of j possible shift values δ , where we seek to estimate the counterfactual mean under each value of δ .
- With this approach, we construct j estimates $\psi_{n,j}$ of the counterfactual mean, each under a different proposed value of the shift δ_j .
- We may summarize $\psi_{n,j}$ through a working marginal structural model (MSM), constructing inference through a hypothesis test of the a parameter of the MSM.
- Formally, let $\vec{\psi}_\delta = (\psi_\delta : \delta)$ with corresponding estimators $\vec{\psi}_{n,\delta} = (\psi_{n,\delta} : \delta)$. Further, let $\beta(\vec{\psi}_\delta) = \phi((\psi_\delta : \delta))$

Variable Importance Analysis with MSMs

- For a given MSM $m_\beta(\delta)$, we have that

$$\beta_0 = \operatorname{argmin}_\beta \sum_\delta (\psi_\delta(P_0) - m_\beta(\delta))^2 h(\delta),$$

- This then leads to the following expansion

$$\beta(\vec{\psi}_n) - \beta(\vec{\psi}_0) \approx -\frac{d}{d\beta} u(\beta_0, \vec{\psi}_0)^{-1} \frac{d}{d\psi} u(\beta_0, \psi_0)(\vec{\psi}_n - \vec{\psi}_0),$$

- In terms of the efficient influence function (EIF) of ψ by using the first order approximation

$$(\psi_n - \psi_0)(\delta) = \frac{1}{n} \sum_{i=1}^n \operatorname{EIF}_{\psi_\delta}(O_i), \text{ where } \operatorname{EIF}_{\psi_\delta} \text{ is the efficient influence function (EIF) of } \vec{\psi}$$

Variable Importance Analysis with MSMs

- Now, say, $\vec{\psi} = (\psi(\delta) : \delta)$ is d -dimensional, then we may write the efficient influence function of the MSM parameter β (assuming a linear MSM) as follows

$$\text{EIF}_{\beta}(O) = \left(\sum_{\delta} h(\delta) \frac{d}{d\beta} m_{\beta}(\delta) \frac{d}{d\beta} m_{\beta}(\delta)^t \right)^{-1} \cdot \sum_{\delta} h(\delta) \frac{d}{d\beta} m_{\beta}(\delta) \text{EIF}_{\psi_{\delta}}(O),$$

where the first term is of dimension $d \times d$ and the second term is of dimension $d \times 1$.